



СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ



«ОСНОВИ РОЗПОДІЛЕНИХ ТА ПАРАЛЕЛЬНИХ ОБЧИСЛЕНЬ»

Шифр та назва спеціальності:	122 Комп'ютерні науки та інформаційні технології	Факультет/Інститут	Комп'ютерних наук та програмної інженерії
Назва освітньо-наукової програми	Комп'ютерні науки та інформаційні технології	Кафедра	Інформатика та інтелектуальна власність

ВИКЛАДАЧ



Любченко Наталія Юріївна, nlubchenko63@gmail.com

Кандидат технічних наук, доцент, доцент кафедри інформатики та інтелектуальної власності НТУ «ХПІ». Досвід роботи – 15 років. Автор понад 60 наукових та навчально-методичних праць. Провідний лектор з дисциплін: «Основи програмування», «Інтелектуальний аналіз даних», «Семантичний WEB», «Об'єктно-орієнтоване програмування»

ЗАГАЛЬНА ІНФОРМАЦІЯ ПРО ДИСЦИПЛІНУ

Анотація	Дисципліна забезпечує теоретичну і практичну підготовку в області паралельних та розподілених обчислень, оволодіння концепціями сучасного програмування в рамках парадигм паралельного та розподіленого програмування. Основу вивчення складають підходи до програмування в багатопотокових системах, розподілених системах, системах синхронних паралельних розрахунків, розглядаються проблеми сумісної роботи процесів паралельної програми та їх синхронізації.
Мета та цілі	Вивчення теоретичних основ та практичних аспектів використання паралельних обчислювальних систем для вирішення складних прикладних задач з великим об'ємом обчислень.
Формат	Лекції лабораторні роботи, консультації. Підсумковий контроль – іспит
Результати навчання	Знати основні методи і засоби паралельної та розподіленої обробки інформації: алгоритми паралельної та розподіленої обробки, засоби їх представлення, засоби програмування паралельних та розподілених обчислень та їх реалізації; уміти застосовувати набуті знання при розробці алгоритмів, апаратних та програмних засобів паралельної та розподіленої обробки.
Обсяг	Загальний обсяг дисципліни 120 год.: лекції – 32 год., лабораторні роботи – 32 год., самостійна робота – 56 год.
Пререквізити	«Основи програмування», «Об'єктно-орієнтоване програмування»
Вимоги викладача	Студент зобов'язаний відвідувати всі заняття згідно розкладу. Дотримуватися етики поведінки. Для проходження дисципліни необхідно мати: ноутбук, лабораторний журнал. Працювати з навчальної та додатковою літературою, з літературою на електронних носіях і в Інтернеті. При пропуску лекційних занять проводиться усна співбесіда за темою. Відпрацювати

лабораторні заняття при наявності допуску викладача. З метою оволодіння необхідною якістю освіти з дисципліни потрібно відвідуваність і регулярна підготовленість до занять. Без особистої присутності студента підсумковий контроль не проводиться.

СТРУКТУРА ДИСЦИПЛІНИ

Лекція 1	Поняття про паралельні та розподілені обчислення	Лабораторна робота №1	Знайомство з основними сервісами Apache Hadoop	Самостійна робота	Система Hive – SQL-інтерфейс доступу до даних платформи Apache Hadoop
Лекція 2	Введення в розподілені розрахунки. Модель, спеціалізації розподілених розрахунків	Лабораторна робота №2	Рішення задачі запису даних в HDFS. Особливості управління ресурсами в режимі Hadoop кластера з YARN		Система Impala – аналіз даних, що зберігаються в розподілених файлових системах
Лекція 3	Програмування з роздільними змінними. Процеси та синхронізація	Лабораторна робота №3	Знайомство з Hadoop MapReduce		Порівняння технологій MapReduce: Hadoop і Twister
Лекція 4	Розподілене програмування. Моделі взаємодії процесів	Лабораторна робота №4	MapReduce, додаткові функції. Ланцюжки MapReduce завдань		Система Mahout – реалізація розподілених алгоритмів ML, орієнтованих на лінійну алгебру
Лекція 5	Розподілений аналіз даних і кластерні обчислення. Розподілені обчислення в задачах машинного навчання	Лабораторна робота №5	Мова програмування Scala. Ознайомлення з синтаксисом і основними можливостями мови Scala		Файлова система MapR
Лекція 6	Мета та задачі паралельної обробки даних. Принципи розробки паралельних методів	Лабораторна робота №6	Знайомство з Apache Spark. Архітектура Spark. Варіанти запуску Spark- додатків		Системи планування завдань
Лекція 7	Моделі обчислень та методи аналізу ефективності	Лабораторна робота №7	Конфігурація і системні утиліти Apache Spark, взаємодія з класичними та розподіленими файловими системами		Система Mesos – система управління кластером для розподілених комп'ютерних середовищ
Лекція 8	Технологія OpenMP для розробки паралельних програм для систем із спільною пам'яттю	Лабораторна робота №8	Використання різних форматів зберігання даних і використання Spark RDD, dataframes і datasets		Подібність та розбіжність MPI, OpenMP, IntelTBB, TPL та MC#
Лекція 9	Розробка паралельних програм для систем з розподіленою пам'яттю з використанням бібліотеки MPI	Лабораторна робота №9	Machine Learning в Apache Spark		Алгоритми паралельного розв'язку диференціальних рівнянь
Лекція 10	Паралельні алгоритми розв'язку задач.	Лабораторна робота №10	Streaming в Apache Spark		

ЛІТЕРАТУРА ТА НАВЧАЛЬНІ МАТЕРІАЛИ

1. Аксак Н.Г. Паралельні та розподілені обчислення: підручник / Н.Г. Аксак, О.Г. Руденко, А.М. Гуржій. – Х.: Компанія СМІТ, 2009. – 480с.
2. Wiktorski Tomasz. Data-intensive Systems: Principles and Fundamentals using Hadoop and Spark. Springer, 2019. — 105 p.
3. Є.Ваврук, О.Лашко Організація паралельних обчислень// Навчальний посібник з дисципліни «Паралельні та розподілені обчислення» для студентів базового напрямку 6.0915 — «Комп'ютерна інженерія» – Л.: Вид. Нац. ун-ту «Львівська політехніка», 2007. – 70 с.
4. Chuck L. Hadoop in Action. – Manning; 1st edition (November 30, 2010) – 325 p.
5. Karau H., Konwinski A., Wendell P., Zaharia M. Learning Spark: Lightning-Fast Big Data Analysis. – O'Reilly Media, 2015. ISBN-13 : 978-1449358624. – 276 p.

1. Horstmann C.S. Scala for the Impatient. – Addison-Wesley. ISBN: 978-0-13-454056-6, 2017. – 369 p.
2. Проект Apache Hadoop [Електронний ресурс]. – Режим доступу: <https://hadoop.apache.org/>
3. Tutorial on Hadoop HDFS and MapReduce [Електронний ресурс]. – Режим доступу: http://hortonworks.com/wp-content/uploads/2012/03/Tutorial_Hadoop_HDFS_MapReduce.pdf
4. Apache Spark Machine Learning Tutorial [Електронний ресурс]. – Режим доступу: <https://mapr.com/blog/apache-spark-machine-learning-tutorial/>
5. Ерл Т. Основи Big Data: концепції, алгоритми і технології. – К: Баланс бизнес букс, 2018. – 320 с.

ПЕРЕЛІК ЗАПИТАНЬ ДЛЯ ПІДГОТОВКИ ДО ІСПИТУ

Поняття про паралельні та розподілені обчислення. Актуальність використання паралельних обчислень. Сфери застосування паралельних обчислень. Вимоги до розподілених систем: прозорість, відкритість, масштабованість. Складність розробки розподілених систем. Програмна реалізація синхронізованої багато поточності. Блокування та бар'єри. Критичні секції. Бар'єрна синхронізація. Семафори, розподілення ресурсів планування. Монітори, методи синхронізації. Особливості програмування під архітектуру з розподіленою пам'яттю. Розгляд задачі розподілення обчислень великих даних. Проблеми використання паралелізму. Типові області використання паралельної обробки даних. Моделювання паралельних програм. Методика розробки паралельних алгоритмів. Залежність ефективності паралельних обчислень від особливостей апаратури. Загальна характеристика стандарту OpenMP. Система MPI. Загальна характеристика. Паралельні алгоритми.

ПЕРЕЛІК ОБЛАДНАННЯ

Лабораторний практикум укомплектовано наступним устаткуванням: комп'ютери з операційною системою Windows -10 з підключенням до мережі Internet.

СИСТЕМА ОЦІНЮВАННЯ

Розподіл балів для оцінювання успішності аспірантів	Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	Нарахування балів
	90-100	A	відмінно	
	82-89	B	добре	
	74-81	C		
	64-73	D	задовільно	
	60-63	E		
	35-59	FX		
	0-34	F	незадовільно з обов'язковим повторнім вивченням дисципліни	

Бали нараховуються за наступним співвідношенням:

- лабораторні роботи: 20% семестрової оцінки;
- самостійна робота: 20% семестрової оцінки;
- іспит: 60% семестрової оцінки

НОРМИ АКАДЕМІЧНОЇ ЕТИКИ

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при нерозв'язності конфлікту доводиться до завідувача кафедри.

Силабус за змістом повністю відповідає робочій програмі дисципліни